

Skatteverkets modell för uppskattning av folkbokföringsfelet: modellspecifikation i årsredovisningen 2019

Innehåll

1	Introduktion.....	3
2	Övertäckningsfel.....	3
3	Undertäckningsfel	7
4	Felaktig folkbokföringsadress.....	7
5	Resultat	9
	Referenser	9

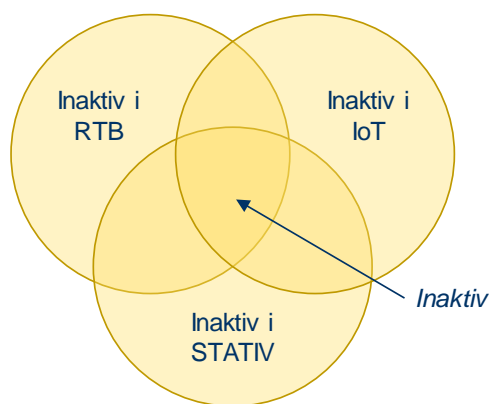
1 Introduktion

Denna promemoria beskriver hur de modellbaserade skattningarna av folkbokföringsfelet i Skatteverkets årsredovisning 2019 har tagits fram. Modellspecifikationerna bygger på ett metodarbete som startade 2018 i och med att Skatteverket fick i regeringsuppdrag att utveckla en metod för att systematiskt kunna följa upp kvaliteten i folkbokföringsregistret. Uppdraget redovisades till Finansdepartementet i september 2018 i form av en rapport (Skatteverket 2018). I rapporten redogörs för den övergripande inriktningen på metoden för uppskattning av förekomsten av fel i folkbokföringsregistret. Metoden omfattar uppskattning av tre olika typer av fel: *övertäckningsfel* (som innebär att det finns individer som är folkbokförda trots att de enligt gällande lagstiftning inte ska vara det), *undertäckningsfel* (som innebär att individer inte är folkbokförda trots att de enligt gällande lagstiftning ska vara det) och *felaktig folkbokföringsadress*. Metoderna har utvecklats i samarbete med Statistiska Centralbyrån (SCB) och innebär i korthet att flera individregister används för att skapa indikatorer på att en viss typ av folkbokföringsfel kan förknippas med vissa företeelser på individnivå. Metodarbetet presenteras i en underlagsrapport från SCB (SCB 2018). Under 2019 gav Skatteverket i uppdrag till SCB att vidareutveckla modellerna i samarbete med Skatteverket. Detta arbete resulterade i de slutliga modellspecifikationerna som används för att ta fram de resultat som redovisas i Skatteverkets årsredovisning 2019. Samtliga beräkningar har gjorts av SCB på uppdrag av Skatteverket.

Denna promemoria utgör en dokumentation över de slutliga modellspecifikationerna som används i årsredovisningen 2019. Promemorian är avsiktligt kortfattad men ambitionen har varit att texten ska kunna stå på egna ben. För bakgrund, detaljer och olika typer av överväganden hänvisas till Skatteverket (2018) och SCB (2018). En betydande del av beskrivningarna i den efterföljande texten kommer från dessa två rapporter och från avsnittet om folkbokföringsfelet i Skatteverkets årsredovisning 2019.

2 Övertäckningsfel

Modellen för övertäckningsfel innehåller två steg. Först definieras en delpopulation av individer med misstänkt övertäckning. Sedan används registerbaserade indikatorer på individnivå för att skatta den totala förekomsten av övertäckningsfel i denna delpopulation. Indikatorernas skattade förmåga att identifiera övertäckning bygger på statistiska samband. Populationen med misstänkt övertäckning består av individer som saknar aktivitet i SCB-registren RTB, STATIV och IoT. Figur 1 illustrerar principen.



Figur 1 Definition av inaktiv (och därmed tillhörande misstänkt övertäckning)

Dessa register beskrivs i detalj i SCB (2015b). SCB baserar sina uppgifter i RTB (Registret över totalbefolkningen) på den information som finns i Skatteverkets folkbokföringsregister. STATIV är en longitudinell individdatabas som innehåller uppgifter om immigration, demografi, utbildning, med mera. IoT (Inkomst- och taxeringsregistret) ligger till grund för statistiken för inkomster och skatter. De statistiska sambanden mellan indikatorerna och sannolikheten att tillhöra övertäckningen skattas för inkomståret 2010. Modellen används sedan för att få årsvisa skattningar av övertäckningsfelet som bygger på de skattade statistiska sambanden för 2010 ihop med respektive års delpopulation med misstänkt övertäckning med tillhörande indikatorer på individnivå.

De statistiska sambanden mellan indikatorerna och sannolikheten för att en individ tillhör övertäckningen skattas med hjälp av logistisk regression. Den beroende variabeln har värdet 1 om individer tillhör övertäckningen ($\hat{O}_i=1$) och 0 om individen tillhör RTB ($\hat{O}_i=0$). Sannolikheten för övertäckning, som vi kan uttrycka som P_i , varierar inom intervallet 0–1 för individerna i den misstänkta övertäckningen. I denna modell skattas den totala övertäckningen som:

$$n^{\hat{O}} = \sum_{i=1}^n \hat{P}_i$$

där \hat{P}_i anger den skattade sannolikheten att individ i tillhör övertäckningen. Den skattade sannolikheten för respektive individ, \hat{P}_i , utgör i sin tur en funktion av individens indikatorvärden och skattade parametrar för respektive indikator.¹

För att skatta denna modell behövs dels ett antal indikatorer, vilka beskrivs nedan, dels data över varje individs övertäckningsstatus ($\hat{O}_i=1$ respektive $\hat{O}_i=0$). En svårighet i sammanhanget är att övertäckningsstatus, till skillnad från indikatorerna, inte kan observeras direkt i data. Detta hanteras på följande sätt. Urvalet för modellskattningen består av 2010 års misstänkta övertäckning. Individerna i urvalet har studerats under följande år, till och med 2017, för att se om de då blivit avregistrerade från RTB (och därmed bekräftats tillhöra övertäckningen även 2010) eller har visat aktivitet under dessa efterföljande år (och därmed bekräftats tillhöra RTB även 2010). Att skapa ett ”facit” på detta sätt vilar på antagandet att individen hade samma status (övertäckning eller RTB-tillhörighet) 2010 som under det senare år som statusen kan konstateras.

De 24 indikatorerna som ingår i den slutgiltiga modellspecifikationen bygger ursprungligen på en registerstudie av SCB (2015a), där 27 indikatorer används för att mäta övertäckningen med hjälp av en modell som kan kallas för deterministisk. Enkelt uttryckt utgörs sambanden mellan indikatorerna och övertäckning i den studien av förutbestämda vikter, till skillnad från de skattade statistiska sambanden som beskrevs ovan. Förutom denna skillnad mellan SCB (2015a) och modellen som används för Skatteverkets årsredovisning 2019 har några indikatorer tagits bort. Det handlar om indikatorer som kräver framåtblickande data för att kunna beräknas.

Nedan redovisas de 24 indikatorerna som ingår i den slutgiltiga specifikationen:

1. **Gränspendling:** Om inaktivitet sammanfaller med registrerad gränspendling finns det starka skäl att tro att personen tillhör RTB-populationen. Registervariabeln för gränspendling är en mycket tillförlitlig variabel. Denna regel granskar bara det aktuella

¹ Funktionen som tilldelar respektive individ en sannolikhet för övertäckning kan uttryckas som $\hat{P}_i = F(\mathbf{I}_i' \hat{\boldsymbol{\beta}})$, där \mathbf{I}_i är en vektor bestående av individens alla indikatorvärden och $\hat{\boldsymbol{\beta}}$ är en vektor med de skattade parametrarna. Funktionen, F , som tilldelar varje individ sin sannolikhet givet innehållet i indikator- och parametervektorerna, består i detta fall av den logistiska kumulativa fördelningsfunktionen. Parametrarna $\hat{\boldsymbol{\beta}}$ skattas med en logit-modell med övertäckning (\hat{O}_i) som beroende variabel och indikatorerna \mathbf{I}_i som oberoende variabler.

året och fungerar därför lika bra för alla undersökta år. En brist är att det endast finns data för gränspendling till Norge och Danmark.

2. **Högskolestudier:** Om inaktivitet sammanfaller med högskolestudier finns det skäl att tro att personen tillhör RTB-populationen. För att avgöra om högskolestudier pågår måste personen blivit registrerad för mer än 10 högskolepoäng. I praktiken motsvarar den en halv termin med heltidsstudier. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
3. **Återkommande aktiviteter i RTB utan aktivitet i STATIV och IoT:** En person som inte gör avtryck i STATIV eller IoT, men har återkommande RTB-händelser som gör personen aktiv dessa år, kan antas tillhör RTB-populationen även de mellanliggande åren. Personen har bevisat att den tillhör RTB-populationen utan att det syns i STATIV eller IoT och eftersom RTB-händelser är av karaktären sällanhändelser är det troligt att personen tillhör RTB-populationen även mellanliggande år. Den valda regeln kräver data från minst 6 år varav minst 3 ska ha en RTB-händelse.
4. **Inaktiv med aktiv RTB-familjemedlem:** En RTB-familjemedlem är en person som har samma familje-ID som den undersökta personen. Om en person som är inaktiv har någon RTB-familjemedlem som är aktiv har personen ändå en stark koppling till att själv tillhöra RTB-populationen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år. Familje-ID är personnumret för den äldste personen i en RTB-familj.
5. **Inaktiv registrerad skriven på församlingen/kommunen:** En person som är skriven på församlingen/kommunen tyder på att myndigheten vet att personen åtminstone har funnits i församlingen/kommunen, men inte haft någon adress. Personer som registreras på församlingen/kommunen kan mycket väl vara sådana personer som har små eller inga aktiviteter i de aktuella registren. Det är därför troligt att personen finns kvar och att personen tillhör RTB-populationen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
6. **Utländska medborgare som har immigrerat efter 60 års ålder:** En person som immigrerar efter 60 års ålder antas ha tjänat ihop till sin pension i något annat land. Det gör att personen troligtvis inte gör avtryck i IoT eller STATIV och eftersom RTB-händelser inte inträffar så ofta kan personen se inaktiv ut fast den tillhör RTB-populationen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
7. **Inaktiv med ränteinkomster:** En person med ränteinkomster som är två prisbasbelopp eller större antas klara sig utan andra inkomster. Dessutom bör en person med höga ränteinkomster ha en ganska stark anknytning till Sverige. När det finns möjlighet till försörjning och anknytning till Sverige är det en indikation på att personen tillhör RTB-populationen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
8. **Familjeinkomst som är större än 0:** Om det finns någon i RTB-familjen som har en inkomst i Sverige får det fungera som en indikator för att personen tillhör populationen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
9. **Högskolestudier följt av inaktivitet för utländska medborgare:** En utländsk student som kommer till Sverige för att studera ska folkbokföra sig här om studierna är längre än ett år. Om studenten sedan lämnar Sverige och inte anmäler emigration efter avslutade studier blir denne inaktiv. Regeln granskar föregående år, vilket gör att regeln kan missa inaktiva som studerat före år 2000.
10. **Studier som grund för bosättning och inaktiv:** En person som har studier som grund för bosättning och inte finns med i högskoleregistret har troligen avslutat sina studier. Om personen då också är inaktiv är det en indikation på att studierna är klara och

personen har lämnat Sverige. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.

11. **Immigration följt av inaktivitet samt ingen familj i Sverige:** En person som nyligen immigrerat och som inte syns i något register och inte har någon känd familj i Sverige. Normalt syns nyligen immigrerade personer i ett eller flera register. Under dessa omständigheter är det troligt att personen har lämnat Sverige. Regeln granskar föregående år vilket gör att regeln kan missa inaktiva som immigrerat före år 2000.
12. **Immigration följt av aktivitet i IoT följt av inaktivitet:** En person som immigrerar för att arbeta i Sverige blir aktiv i IoT. Om sedan personen lämnar Sverige utan att anmäla emigration kommer den att bli inaktiv. En person som kommit hit för att arbeta under en kort period kan mycket väl sakna kunskap eller intresse att själv anmäla emigration. Regeln granskar två föregående år, vilket gör att regeln kan missa inaktiva som immigrerat före år 2000.
13. **Inaktiv utan känd hemvist:** En person som saknar känd hemvist och samtidigt är inaktiv är en klar indikation på att personen tillhör övertäckningen. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
14. **Inaktivitet och inkomst som är 0 efter att ha varit mer än 0:** En person som blir inaktiv samtidigt som inkomsten blir noll efter att ha haft inkomst tidigare, är en indikation på att personen troligen lämnat Sverige. Regeln granskar föregående år, vilket gör att regeln kan missa inaktiva som haft en inkomst före år 2000.
15. **Utländsk medborgare som blir ensam i sin familj:** En utrikes född person eller utländsk medborgare som får förändrade familjeförhållanden och blir ensam i familjen och samtidigt blir inaktiv, är en indikation på att personen har lämnat Sverige. Regeln granskar föregående år, vilket gör att regeln kan missa inaktiva som haft andra familjeförhållanden före år 2000.
16. **En yngre familjemedlem finns kvar på samma adress:** Om en person som tillhör ett familje-ID och detta byts till den aktuella personens familje-ID betyder det att de äldre personerna i familjen har flyttat. Om en familj splittras och en yngre person blir kvar på samma adress samtidigt som den blir inaktiv är det en indikation på övertäckning. Regeln granskar föregående år, vilket gör att regeln kan missa inaktiva som haft andra familje-ID före år 2000.
17. **Inaktiv med arbete som grund för bosättning:** En person som haft arbete som grund för bosättning och blir inaktiv bör kunna ses som en indikation på övertäckning. Regeln granskar bara det aktuella året och fungerar därför lika bra för alla undersökta år.
18. **Inaktiv efter död:** En person som är död bör inte ingå i RTB-populationen. Den här regeln ska inte få några träffar, men om så är fallet bör det vara frågan om övertäckning. Regeln granskar föregående år vilket gör att regeln kan missa inaktiva som har markerats döda före år 2000.
19. **Endast garantipension:** Även personer som inte bor i Sverige kan få garantipension. Därför är inte garantipension i sig en indikator på att personen tillhör RTB-populationen. Om personen inte har någon annan inkomst eller bidrag är det snarare en indikator på att personen tillhör övertäckningen.
20. **Automatisk nolltaxerare över 25 år:** Personer som med automatik blir nolltaxerare har inte lämnat in någon deklaration och därför finns heller inga tecken på att personen finns i Sverige. Barn och ungdomar som går i skolan och inte har några inkomster är normalt automatiska nolltaxerare utan att tillhöra övertäckningen. Är personen däremot över 25 år bör den ha någon inkomst eller annan ersättning (bidrag).
21. **Utgående uppehållstillstånd:** En person som inte har förnyat sitt uppehållstillstånd utan låtit det gå ut under föregående år finns troligen inte länge kvar i Sverige. Regeln

granskar föregående år, vilket gör att den kan missa inaktiva som har utgående uppehållstillstånd före 2010.

22. ***Ej inskriven vid grundskolan:*** Finns en person i skolåldern inte inskriven vid grundskolan den 15 oktober aktuellt år kan man anta att personen troligtvis inte finns kvar i landet. Regeln granskar innevarande år och fungerar bra för alla år efter 2007.
23. ***Anknytning till arbetsmarknaden:*** Alla personer med någon anknytning till arbetsmarknaden, arbetar, har påbörjat arbete och har slutat arbetat, innebär att det någon gång under innevarande år varit i Sverige och jobbat. Regeln granskar innevarande år och fungerar bra på alla år från och med 2003.
24. ***Särskild postadress till utlandet:*** En person som vill ha sin post från Skatteverket till utlandet bor troligen inte i Sverige. Regeln granskar innevarande år och fungerar bra på alla år.

3 Undertäckningsfel

Eftersom individer som tillhör undertäckningen inte förekommer i individregister i samma omfattning som folkbokförda individer, finns inte förutsättningar att tillämpa samma form av indikatormodell för undertäckningsfelet som för övertäckningsfelet.

Undertäckningsmodellen är därför enklare och innebär att tre grupper med misstänkt undertäckning summeras ihop. Den första gruppen består av individer med samordningsnummer som inte är folkbokförda men som under aktuellt år och året före hade en disponibel inkomst över ett prisbasbelopp. I den andra gruppen finns individer med så kallade sena aviseringar, som innebär aviseringar från Skatteverket avseende invandring och nyfödda som inkommit till SCB 1 februari till och med 31 december aktuellt år och som avser tidigare år. Den sista gruppen avser utländska studenter som inte är folkbokförda. Där ingår individer med samordningsnummer som har varit inskrivna i högskolan under höstterminen året före aktuellt år samt under både vår- och höstterminen under aktuellt år.

4 Felaktig folkbokföringsadress

Metoden för att skatta förekomsten av felaktig folkbokföringsadress liknar metoden för skattning av undertäckningen i det att delgrupper med misstänkt felaktig adress identifieras och sedan summeras. Följande fem grupper ingår:

1. ***Studentpendling med långt pendlingsavstånd:*** Här ingår studenter som studerat tre terminer i följd i samma kommun; vår- och hösttermin aktuellt år samt vårterminen året efter aktuellt år. Studierna ska inte bedrivas på distans. Studenterna ska också vara folkbokförda i en annan kommun än studiekommunen 31 december aktuellt år. Fem pendlingsavståndskriterier används:
 - ≥ 40 km
 - ≥ 70 km
 - ≥ 100 km
 - ≥ 130 km
 - ≥ 150 km
2. ***Arbetspendling med långt pendlingsavstånd:*** I gruppen ingår personer som klassas som förvärvsarbetande enligt RAMS² och som arbetar i en annan kommun än där de är

² Se Skatteverket (2015b) för en beskrivning av detta register.

folkbokförda 31 december aktuellt år. Personerna ska inte heller vara gifta eller sambo. Fem pendlingsavståndskriterier används:

- ≥ 400 km
- ≥ 500 km
- ≥ 600 km
- ≥ 700 km
- ≥ 800 km

3. **Personer som bor på liten boarea:** Här ingår personer som 31 december aktuellt år är folkbokförda i en lägenhet där boarean per person understiger 9 m^2 . Fem kriterier för boareans storlek används:

- $< 9 \text{ m}^2$
- $< 8 \text{ m}^2$
- $< 7 \text{ m}^2$
- $< 6 \text{ m}^2$
- $< 5 \text{ m}^2$

4. **Personer folkbokförda i stora övriga hushåll, det vill säga stora hushåll med personer utan nära anhörig:** Gruppen består av personer som 31 december aktuellt år är folkbokförda i en lägenhet där det finns minst 3 så kallade övriga personer, dvs. personer som inte har någon nära anhörig i samma hushåll. Följande kriterier för vad som betraktas som ett stort hushåll används:

- ≥ 4 personer i hushållet
- ≥ 6 personer i hushållet
- ≥ 8 personer i hushållet
- ≥ 10 personer i hushållet
- ≥ 12 personer i hushållet

5. **Personer som efter impulser om felaktig adress har fått sin folkbokföringsadress ändrad av Skatteverket:** Grundurvalet består av samtliga inkomna ärenden till Skatteverket (impulser) som handlar om underrättelse om felaktig folkbokföring. En delpopulation av dessa ärenden har resulterat i ändrat folkbokföringsdatum under tiden kontrollärendet var oavslutat eller inom 7 dagar efter avslutat kontrollärende. Det är denna delpopulation som ingår i skattningen.

Eftersom det finns osäkerheter om vad som ska betraktas som långt pendlingsavstånd, liten boarea och stort hushåll bygger gruppindelning 1-4 ovan på genomsnittet av delgrupper definierade enligt de olika kriterierna.³ Den slutliga sammanräkningen av individer som ingår i grupp 1—5 redovisas i nettotermer, det vill säga siffrorna mäter antalet individer med felaktig folkbokföringsadress och inte antalet indikationer på felaktig adress enligt kriterierna i någon av grupperna.⁴

³ Se fotnot 28 i Skatteverket (2018) för en diskussion om vad det innebär att ta ett genomsnitt i det här fallet.

⁴ Ordningen i vilken dubletter tas bort påverkar det redovisade antalet individer i varje enskild grupp, men ordningen spelar ingen roll för den totala nettosumman. Se SCB (2018) för mer detaljer.

5 Resultat

Nedan redovisas resultatet från tillämpning av modellerna för övertäckning, undertäckning och felaktig folkbokföringsadress som beskrevs ovan. Tabellens innehåll är densamma som i Skatteverkets årsredovisning 2019.

Tabell Uppskattning av folkbokföringsfelet (antal individer i 1000-tal)

Antal	2012	2013	2014	2015	2016	2017
Övertäckning	65,6	67,3	67,1	61,5	62,2	64,6
Undertäckning						
Utländska studenter	1,0	1,1	1,1	0,9	1,0	1,1
Ej folkbokförda med samordningsnummer och inkomster under två efterföljande år	3,4	3,7	4,2	4,8	5,6	7,5
Sena aviseringar	1,1	1,3	1,4	1,9	1,3	3,2
Summa	5,5	6,0	6,7	7,7	7,9	11,8
Felaktig adress						
Studentpendling	10,9	11,2	11,2	11,4	11,5	11,5
Arbetspendling	23,5	23,2	24,0	24,3	24,2	24,6
Liten boarea	32,0	35,1	36,2	36,3	40,1	40,9
Stora hushåll	11,7	12,7	13,9	15,0	17,7	18,3
Impulser	27,5	33,0	30,2	26,0	28,4	23,6
Summa	105,7	115,2	115,5	113,1	122,0	119,0
Totalsumma	176,7	188,5	189,3	182,2	192,1	195,4
Andel av befolkningen	1,8%	2,0%	1,9%	1,8%	1,9%	1,9%

Källor: SCB:s beräkningar av Skatteverkets modell för skattning av folkbokföringsfelet samt egna beräkningar med hjälp av årlig befolkningsstatistik från SCB.

Referenser

SCB (2015a), *Övertäckning i Registret över totalbefolkningen – en registerstudie*, Befolkning och välfärd 2015:1.

SCB (2015b), *SCB:s data för forskning – de mest använda registren*.

SCB (2018), *Folkbokföringsfelet: en metodrapport om skattning av övertäckning, undertäckning och folkbokförda på fel adress*, tillgänglig på www.skatteverket.se.

Skatteverket (2018), *Kvalitetsuppföljning i folkbokföringsregistret*, tillgänglig på www.skatteverket.se.